

Analysis of Machine Learning Algorithms for Predicting Depression

MS. Purude Vaishali Narayanrao
Research Scholar,
Department of Computer Science And Engineering,
Koneru Lakshaiah University (KLEF),
Hyderabad, Telangana, India.

Dr. P. Lalitha Surya Kumari
Professor,
Department of Computer Science And Engineering,
Koneru Lakshaiah University (KLEF),
Hyderabad, Telangana, India.

Abstract — The increasing use of technology can cause a life- style of less physical work. Also, the constant pressure on an individual can create a risk of mental disorder. These vulnerabilities include peer pressure, heart attack, depression, and many other effects. In this paper different approaches to predict depression are studied in detail. The mechanisms include collecting dataset through questionnaires asked to the person, posts on social media, text used throughout verbal communication and expressions on face. Result is derived from extracted information. Here output expected is that the person needs attention or not. In this research work different algorithms and classifiers of machine learning such as Decision Trees , SVM, Naive Bayes Classifier, Logistic Regression and KNN Classifier are analyzed to identify state of mental health in a target group. Target groups used for this identification process are public like students of high school, college students and working professionals. The paper also demonstrates an example in which Twitter scrapping tool Twint is used detect whether given Twit is depressive or not.

Index Terms—Machine Learning, Mental health, Predictive Analytics.

I. INTRODUCTION

As per WHO, 4.4% on globe are suffering with depression, among that varieties of depression are more common in female than male. A major reason for suicide and global disability is Depression. If not treated, it affects day-to-day life of people surrounding the person who is in depression, in family, at workplace or in society. Depression causes negative thinking, less concentration in work, reduced productivity. It also affects reproduction system of human being. Depression may cause mental disorder also. Early detection of mental disorder helps to start treatment at earlier stage and improves quality of patient's life as well as his family. Mental health handles stress and is a measure for taking decisions in each step in life. Mental Health is very important factor in every stage in life whether it be childhood or an adult. Mental health could affect one's thinking and behavior. Positive mental health will support person to have full potential. It will also help him to cope up with the stress at home and workplace. It will increase the productivity of the people.

II. METHODOLOGY

For predicting the Mental illness the algorithms of Machine learning and Artificial Intelligence (AI) are very much useful. In real time these applications are helpful for the society as it will serve as a monitoring tool for

individuals having different behavior. Prediction is performed using Predictive Analytics.

Along with statistics (both historical and current), Machine Learning algorithms, Predictive modeling and Data mining techniques are used to estimate, or predict future outcomes. Predictive Analytics is a branch of analytics that forecast any future activity, behavior and trends which is called prediction.

Typically, this is the field of data scientists, statisticians and skilled data analysts. Data relevant to this task is gathered by Data engineers & then prepare it for analysis and also help required for data visualization, dashboards and reports is provided by software developers and business analysts. Data scientists use Predictive models to find correlations between different data elements in stream data, patient health records and other types of data sets. After collection of the dataset , a statistical model is developed, trained and changed as required to get correct results. Against the selected data the model is then run to generate predictions.

III. RELATED WORK

Related work useful for this research is studied in detail. In all studies different inputs are facial expression, Video, Text, Posts & comments on social media & behavioral features. Videos of population are captured & used as input for detecting depression [1]. Facial expressions from captured video are identified. But expression changes depending upon pose, lighting, person, angle and sensors. So it's a challenging task to encode features to detect depression. Currently Deep learning architectures are providing good performance. For two dimensional & three dimensional Convolutional Neural Network & Recurrent Neural Network, Spatial and temporal information is used. In which a deep learning architecture is proposed to find the ordinal relationship between the expressions on face and intensity of depression based on distribution learning. Method depends on a new expectation loss. Thereby performance of the model is increased without the need to employ multiple channels. Performance is evaluated using MAE and RMSE. When a video is inputted, score is calculated for each individual frame of video. All these scores are averaged & based on it overall predicted depression score is calculated. The average pooling efficiently summarize spatio-temporal features. Hence author in [2] has detected that fully-connected layers are not required at all and it has also caused decrease in the number of model parameters required along with over-fitting.

In their experiments it is proved that when we combine global & local responses, it shows increase in performance and accuracy of C3D networks. While author in [6] proposes a system that will help in detecting depression in college students. Features of happy, contempt and disgust faces are used to train the system. For testing phase college students are asked different questionnaires and videos of students are collected while they are answering different questionnaires. For effective detection of features throughout the video, the facial features of these students are extracted and normalized. For the test dataset then the extraction of facial features would be done, classifying them by SVM classifier for depression detection. In the training phase the SVM classifier is trained with happy, contempt and disgust features extracted from the input happy face dataset. With this trained SVM the features extracted from each frame of the video dataset can be tested to check if these features are present in each frame. An indication about the level of depression is obtained from the level of these features. Author in [14] has inputted Visual & Vocal information using camera & microphone. These 2 inputs are used as keys for developing strong system based on AI for recognizing depression levels.

Using pertained CNNs artificial image processing can be exploited by using the visual modality. For example VGGFace provide good features at the frame level of videos. But it can only work for still images. Hence to capture spatial as well as dynamic temporal information technique called Feature dynamic history histogram (FDHH) technique is proposed. ML algorithms like PLS & LR are used to map dynamic features and the depression scales.

It is noticed that depression and other mental illnesses may lead to social withdrawal and isolation. As it is observed that social media is used by everyone to share happiness, joy, achievements as well as sadness. Depressed people will use social media for connecting with other people, for sharing joy,happiness,sadness experience & for supporting each other [3]. They are expecting professional help online which can be provided using social media [13].

To identify & divide depression symptoms from written texts in early stage, there are different ways based on machine learning methods. When a person is affected by depression the language person speaks will definitely changes. Mental health can be monitored using count of specific words by maintaining dictionary of such words. The developed software is Linguistic Inquiry and Word Count (LIWC) [3]. For same purpose a Python Library Differential Language Analysis Toolkit (DLATK) was created which is open-source [3].

In [3] dataset is formed by continuously monitoring posts & comments on a site reddit.com. It has a group of 752 users out of which 135 are depressed. Posts are searched with a message which mention either depression or treatment (e.g. "I undergone for treatment of depression"). If match is found then that user is identified as depressed user. Author in this paper invented ERDE measure for early detection of risk of depression. To represent the best

possible ERDEo the prediction thresholds have been chosen.

In [4] a system based on AI is proposed which classifies users according to their mental health. Input used for this system is post made by user on Social Networking site [SNS]. System scans user generated content (UGC) in the post. Classifiers used are : SVM and Naïve Bayes. The performance is analyzed using accuracy, precision and recall. The accuracy means how much percentage of depression is predicted correctly, the precision means how many predictions are positive (Non-Depressed), and the recall denotes the percent of the positive (Not- Depressed) correct cases caught. The experimental results are achieved by testing 30 people. Out of 30 people, 15 are detected depressive after answering the BDI-II questionnaire, and the remained are not predicted as depressed.

In [8] previous work is reviewed and the overall phases of MHMS(Mental Health Monitoring Systems) are presented by proposing a Classification Taxonomy.

Different classification algorithms & clustering techniques are used to identify the mental health of an individual in a population. A prediction model is built for assessing the mental health. In this research work clustering is used to identify the possible groups in the two target population taken under consideration. These are named as population1 & population2. For each group 300 & 356 samples were collected respectively. Then various clustering algorithms like K-Means clustering, hierarchical clustering and partitioning around medoids or K- medoids are applied. The techniques SVM,KNN, ensemble (bagging) and tree ensemble (random forest) give an equivalent accuracy score of 0.9. In order to validate the labels obtained by clustering the concept of MOS are used.

There is Correlation between depressive symptoms in patients and behavioral features [17]. In this research work behavioral features are collected using mobile and wearable devices. For experimental result 46 persons are observed by author and identified 85 features which are unique and shows different behavior. This all data is collected using 17 sensors. Then identified features were divided into 7 categories. Out of 85 some features are found having major and consistent correlation that is directionality with mood assessment (e.g. the amount of home stay, sleep duration, and vigorous activity), while others showed directionality discrepancies across the studies(e.g. amount of text messages [SMS] sent, time spent between locations, and frequency of phone screen activity).

IV. EXAMPLE

As a example, Dataset collected from Data Driven Investor (DDI) is used which is a small team with the mission to develop a guide from confusion to clarity in the age of information, knowledge, and complexity[21]. It detects depression and mental disorder using tweets from social media Twitter by using algorithms of Machine Learning. It used Baye's classifier & python. The output is whether person is depressed or not.

Keywords are extracted from tweets using tool Twint and depression is detected.

```

8148 When both people in a relationship have mental...
8221 I'm entering a post-infinity war depression bye
8242 @kanyewest @Da Goat21 @Rosie is a true advocat...
8433 Recently fallen off waking up and going for a ...
8549 What! Your #depression is back? Here are some ...
8562 #selegiline dosage for depression https://goo...
8638 Mom's depression tied to kids' emotional, inte...
8685 @hamrick_krista @itsAshleyBB Well Geez Ashley...
8748 i also have a lot to say about asian immigrant...
8750 I want glittery, fun, sexy, crazy party music ...
8775 Mom's depression tied to kids' emotional, inte...
8795 I was supposed to revise business but I forgot...
9007 @ethomasson is a senior correspondent @Reuters...
9062 @LaurakBuzz sadly not as deep as i thought. Bu...
9111 Worst thing about depression is forgetting to ...
9150 Exercising regularly cuts risk of depression b...
9256 But I know depression and drug addiction don't...
9375 @gabrielflorin01 Sis what happened? Depression...
9377 so glad winter is finally over. Goodbye season...
9381 @BethFratesMD That's what I've been trying to ...
9415 My depression won't let me work out and be hea...
9622 Why your doctor doesn't provide #IVKetamine as...
9643 How exercising can slash the risk of depressio...
9732 Going back and listening to an older album of a...
9750 @WhatTheFFacts I don't like to say that I'm go...
9763 Study finds exercise may lower risk of depress...
9779 Poor recognition of #depression and #anxiety i...
9812 Depression is real
9842 SIP works in Flat, Volatile, Bullish , Bearis...
9871 Learning how to forgive myself for the stagnan...

```

B. Positive Tweets

```

7 23 or 24ï¿½ possible today. Nice
10 Good morning everybody!
15 @blueaero ooo it's fantasy? i like fantasy no...
30 got the magic numbers and is thanking y'all fo...
61 Yaaaaay. Jac is training my sorcerer to level ...
62 My doggy is so cute.
103 The official result came out, PR won 2 seats w...
216 New Osalto blog post - http://tinyurl.com/dnbr...
257 @brob108 good stuff Woop, looking forward to...
277 @em_c Good try Well done with the painting by...
288 @heath7s LOLLOL i love a man in uggs.
291 This night It was awesome, I slept at 1 a.m ....
327 @etrangle http://twitpic.com/318mv - Mysha &am...
358 @jonthanjay awesome i have one...there's some...
466 new to twitter world, thinking ob how to best ...
518 i have had a pretty good day today. went to t...
539 @prernabhagna Welcome to the Twitterverse Prerna
556 fare thee well, tweetdom
612 Just gonna have me a good time and screw every...
624 @Golddyfinch too late. hv already crossed 2 the...
632 @winson Great, I look forward to your feedback
808 @sophiestication Good luck Sophia!
898 Too much running around today but at least it ...
922 @anima Holy shit
925 @henriqoliv182 You can't say anything more tha...
984 Fancy is a 187 hun got u doin things u aint ne...
1055 @Juern der maaaaster is back
1084 @dean2105 Stress-free is the way to be
1154 @Skiingfreak77 and by the way it's short stack...
1164 @ElvenstarArt Great !!! Your book looks fantas...

```

As a cause of disability Depression has been recognized in the whole world as it is causing more considerable cost to health care systems. Depression leads to negative thinking as well as insecure feeling. It affects physical well-being and behaviors of the depressed person. In more cases, depression is found as one of the leading causes of suicide and substance abuse. In this paper different approaches used for Predicting Depression are analyzed and a Deep Learning mechanism is proposed for automated depression detection. It allows to extract features from collected dataset thorough text on Twitter. This mechanism has potential to improve accuracy.

VI. REFERENCES

- [1] Wheidima Carneiro de Melo, Eric Granger†, Abdenour Hadid "Depression detection based on Deep Distribution Learning", ICIP 2019, Sep 2019, IEEE.
- [2] Wheidima Carneiro de Melo, Eric Granger†, Abdenour Hadid "Combining Global and Local Convolutional 3D Networks for Detecting Depression from Facial Expressions", IEEE FG, May 2019, IEEE.
- [3] Marcel Trotzek, Sven Koitka, and Christoph M. Friedrich "Utilizing Neural Networks and Linguistic Metadata for Early detection of depression Indications in Text Sequences", IEEE TKDE, Dec 2018.
- [4] Maryam Mohammed , Hafiz Farooq Ahmed "Predicting depression levels using Social Media Posts", Jun 2017, IEEE.
- [5] Sandhya P, Mahek Kantesaria "Prediction of Mental disorder for employees in IT industry", IJITEE Apr 2019.
- [6] Dr. D. Venkataraman , Namboodiri Sandhya Parameswaran "Extraction of Facial Features for Depression Detection among Students", 2018, IJPAM.
- [7] Ms.Sumathi M.R., Dr. B. Poorna "Prediction of Mental Health Problems among Children using Machine Learning Techniques ", IJACSA 2016.
- [8] Enrique Garcia-Ceja , Michael Riegler , Tine Nordgreen ,Petter Jakobsen, Ketil J. Oedegaard , Jim Tørresen "A Mental Health monitoring with Multimodal Sensing & Machine Learning", Science Direct, Sept.2018.
- [9] Petra Hoffmannová "Text-Based Detection of the Risk of Depression , Frontiers in Psychology", March 2019.
- [10] Anastasia Pampouchidou, Panagiotis Simos, Kostas Marias, Fabrice Meriaudeau, Fan Yang, Matthew Padiaditis, and Manolis Tsiknakis "Automatic Assessment of Depression Based on Visual Cues: A Systematic Review", IEEE Transactions on Affective Computing, 2017.
- [11] M. Srividya "Behavioral Modeling for Mental Health using Machine Learning Algorithms", Springer Nature 2018.
- [12] Amir Hossein Yazdavar, Hussein S. Al-Olimat, Monireh Ebrahimi, Goonmeet Bajaj, Tanvi Banerjee, Krishnaprasad Thirunarayan, Jyotishman Pathak, and Amit Sheth "Semi-Supervised Approach to Monitoring Clinical Depressive Symptoms in Social Media", IEEE Conference 2018 April 24.
- [13] "A Framework for Classifying Online Mental Health-Related Communities With an Interest in Depression", IEEE J-BHI, VOL. 20, NO. 4, JULY 2016.
- [14] Budhaditya Saha, Thin Nguyen, Dinh Phung, and Svetha Venkatesh "Artificial Intelligent System for

Automatic Depression Level Analysis Through Visual and Vocal Expressions” , IEEE TCDS, VOL. 10, NO. 3, SEPTEMBER 2018.

[15] Daniel Durstewitz¹, Georgia Koppe, Andreas Meyer-Lindenberg “Deep neural networks in psychiatry”, Springer Nature Feb 2019.

[16]Asma Ghandeharioun, Szymon Fedor, Lisa Sangermano, Dawn Ionescu, Jonathan Alpert, Chelsea Dale, David Sontag, and Rosalind Picard “Objective Assessment of Depressive Symptoms with Machine Learning and Wearable Sensors Data”, 2017, ACII.

[17] Darius A Rohani, Maria Faurholt-Jepsen, Lars Vedel Kessing, Jakob E Bardram “Correlations Between Objective Behavioral Features Collected From Mobile and Wearable Devices and Depressive Mood Symptoms in Patients With Affective Disorders: Systematic Review”,JMIR Mhealth Uhealth 2018.

[18] Khari, M., & Karar, A. (2013), “Analysis on intrusion detection by machine learning techniques: a review”, Int. J. Adv. Res. Computer. Sci. Software Eng, 3(4).

[19]Pritam, N., Khari, M., Kumar, R., Jha, S., Priyadarshini, I., Abdel-Basset, M., & Long, H. V. (2019). “Assessment of code smell for predicting class change proneness using machine learning”, IEEE Access, 7, 37414-37425.

[20]Khari, M., & Kumar, P. (2018). “Evolutionary computation-based techniques over multiple data sets: an empirical assessment”, Arabian Journal for Science and Engineering, 43(8), 3875-3885.

[21] Viridiana Romero Martinez “A Machine Learning Approach for the Detection of Depression and Mental Illness in Twitter”, DDI, Feb 1, 2019

First Author is Asst. Prof. in the CSE Department of NGIT college, Uppal, Hyderabad. She is research scholar in the Department of Computer Science and Engineering, Koneru Lakshaiah University, Hyderabad, Telangana, India. Her research interest includes Image Processing & Artificial Intelligence.

Second Author is Professor in the CSE Department, Koneru Lakshaiah University, Hyderabad, Telangana, India. Author completed Ph.D in CSE department from JNTUH, Hyderabad, Telangana, India. Author is interested for doing research in areas which include N/W Security, Information Security and Algorithms.